ON THE VARIATE DIFFERENCE CORRELATION METHOD AND CURVE-FITTING.

By Warren M. Persons, *Harvard University.*

(Professors A. A. Young of Cornell University and E. E. Day of Harvard University have generously read the manuscript of this paper and have made helpful suggestions. Computations and pertinent comments have been made by Mr. Edwin Frickey of Colorado College. W. M. P.)

I.

In *Biometrika* for April, 1914, "Student" generalized the method of differences for the elimination of spurious correlation due to order of items in time or space. In the same *Journal* for November, 1914, Dr. O. Anderson of Petrograd provided the probable errors of the successive difference correlations of two series where the correlations of random pairs of the variates are zero. In the same number B. M. Cave and Karl Pearson presented "Numerical Illustrations of the Variate Difference Correlation Method," using Italian economic data.

The conclusion of "Student" is this: "If we wish to eliminate variability due to position in time or space and to determine whether there is any correlation between the residual variations, all that has to be done is to correlate the 1st, 2nd, 3rd. . . nth differences between successive values of our variable with the 1st, 2nd, 3rd . . . nth differences between successive values of the other variable. When the correlation between the two nth differences is equal to that between the two $(n+1)$th differences, this value gives the correlation required."* The meaning of "Student" is that the correlation required is indicated by the ultimate steadiness of values of the correlation coefficient for higher multiple differences of the items.†

After working with 11 series of 28 items each Miss Cave and Professor Pearson stated:‡ "In most cases our difference cor-

* *Biometrika*, April, 1914, p. 179.

† *Biometrika*, November, 1914, footnote p. 340.

‡ *Ibid.*, pp. 354–355.

relations have hardly even with the sixth differences reached a steady state. . . . In the great bulk of instances there is still a more or less steady rising or falling appreciable in the difference correlations, and all we can really say is that the final value, the true r_{XY}, will be somewhat greater or less than a given number. From an examination of the actual numerical working of the correlations, it appears to us that the terminal values are in the case of these short series of very great importance. It is further clear that the theory as given by 'Student' depends upon certain equalities which are not fulfilled in practice in short series. We await with much interest the complete publication of Dr. Anderson's work, and hope to find a fuller discussion of the allowance to be made in short series for the influence of the terminal state of affairs on the steadiness of the series and on the approach to the standard deviation formulae. But apart from these lesser points, our present numerical investigation has convinced us of the very great value of the new method of Variate Difference Correlations."

In the demonstration leading to his theorem, previously quoted, "Student" stated that "if x_1, x_2, x_3, etc., y_1, y_2, y_3, etc., be corresponding values of the variables x and y, then if x_1, x_2, x_3, etc., y_1, y_2, y_3, etc., are randomly distributed in time and space, it is easy to show that the correlation between the corresponding nth differences is the same as that between x and y."* In the proof of this statement certain assumptions were made as follows: first, that there is a significant correlation between items with similar subscripts, *i e.*, $\Sigma x_K y_K \neq 0$; second, that if the items of each series be paired with the preceding items of the same series the correlation will be zero, *i. e.*, $\Sigma(x_K x_{K+1}) = 0$ and $\Sigma y_K y_{K+1} = 0$; third, that there is no significant correlation between the items of one series and the items of the other when there is a lag in either direction, *i. e.*, $\Sigma(x_K y_{K+1}) = 0$, $\Sigma(x_{K+1} y_K) = 0$; fourth, that the sum of the items of each series is zero, *i. e.*, $\Sigma x_K = 0$, $\Sigma y_K = 0$; fifth, that the time element in any series ordered in time can be expressed as an algebraic function of time (t) of some degree, the nth.

* *Biometrika*, April, 1914, p. 179.

It is my contention that these assumptions are such as cannot be retained in applying the method to the most common types of problems. For instance the pairing of items of two time series is made possible by the position of those items in time either because they occur in the same time interval (concurrent) or in definitely related intervals (lag). Our problem may be, and usually is, not only to determine the correlation but to find what pairings give the maximum correlation. In such case the assumption that only one pairing is significant vitiates the conclusion at the outset. The writers on the variate difference correlation method all assume that "the true r_{XY}" is for pairs concurrent in time.

Let the method of variate differences be applied to two time series artificially constructed. Let series A be made up of successive values of the function $2t+3$ with $+1$ and -2 alternately added to the items. Let series B be made up of the values of successive values of the function t^2+3t+2 with -1 and $+1$ alternately added to the items. We will have the following series and differences:

SERIES A.

t	*Items*	*1st D*	*2nd D*	*3rd D*
0	4	..	..	..
1	3	−1	..	..
2	8	+5	+6	..
3	7	−1	−6	−12
4	12	+5	+6	+12
5	11	−1	−6	−12
		etc.		

SERIES B.

t	*Items*	*1st D*	*2nd D*	*3rd D*
0	1	..	..	..
1	7	+6	..	..
2	11	+4	−2	..
3	21	+10	+6	+8
4	29	+8	−2	−8
5	43	+14	+6	+8
		etc.		

The process of taking differences eliminates the elements due respectively to the first and second degree functions of time. The oscillating elements remain. If concurrent items for the third and higher differences are paired, we have $r_o^{iii} = r_o^{iv} = \quad . \quad . \quad . \quad r_o^{(m)} = -1$. If the items are paired with a lag of one in either direction we have $r_{-1}^{iii} = r_{+1}^{iii} = r_{-1}^{iv} = r_{+1}^{iv} = \quad . \quad . \quad . \quad r_{-1}^{(m)} = r_{+1}^{(m)} = +1$.* In this case the time elements are eliminated as far as they can be by the method of differences and yet the series resulting are not, properly speaking, random in time. It will be of interest to determine if this sort of oscillation occurs in actual data.

II.

I have recently been working with some 21 series of economic statistics for the United States for the period 1879 to 1913, giving 35 items to each series. Application of the variate difference correlation method to such series has forced me to the conclusion that neither the possibilities nor the limitations of the method, when applied to short series, have been appreciated by the writers on the subject.†

The applications of the method by "Student" and by Miss Cave and Professor Pearson are not satisfactory tests: first, because they have applied it merely to items of the same date, thus assuming that the real correlation can exist only for such pairs; and, second, they assume that the correlation indicated by the steadiness of coefficients between higher differences only is significant, the coefficient for first differences not being significant unless it is supported by steadiness of the coefficients of higher differences. My objections to these assumptions and the conclusions based upon them will be il-

* In r_i the subscript i indicates the number of time intervals that the items of series A precede (−) or lag behind (+) the corresponding items of series B.

† At the conclusion of his article in *Biometrika*, April, 1914, pp. 269–279, Mr. O. Anderson makes the following statement, which I translate:

"If we take into consideration that for our purposes the evolutionary component of a series has disappeared if it becomes so small relative to the oscillatory component that it can influence only the 3rd, 4th, etc., decimal place of the expression for R [the coefficient of correlation] then we may conclude that not only components which are represented by a parabola of higher order, but also those represented by transcendental functions (such as a sine curve) become eliminated taking by a finite number of differences. Further it may be shown that generally all more or less 'smooth series' all of which are characterized by a considerable degree of positive correlation between adjacent items, lose the character of smoothness in the process of multiple differences. The generalized Cave-Hooker procedure is, therefore, manifestly a quite universal means of sifting out the correlation between the oscillatory elements of complex series."

lustrated by application of the variate difference correlation method to American data for the period 1879–1913. The series used are the following:

1. Wholesale prices of commodities.
2. Gross receipts of railroads.
3. Net earnings of railroads.
4. Coal production.
5. Exports from the United States.
6. Imports into the United States.
7. Pig-iron production.
8. Price of pig-iron.
9. Immigration (fiscal year).
10. Shares sold on the New York Stock Exchange.
11. Average price of shares sold on the New York Stock Exchange.
12. New York clearings plus five times outside clearings (called clearing index).
13. Clearing index divided by relative wholesale prices (called corrected clearing index).
14. New railroad mileage constructed.
15. Per cent. of business failures.
16. Liabilities of business failures.
17. Balance of trade.
18. Weighted index numbers of the yield per acre of nine leading crops.
19. Ratio of loans to resources of banks.
20. Ratio of cash to deposits of banks.
21. Surplus reserves of New York associated banks.

In each case, except for the clearing index, I assumed the secular trend to be linear. A straight line was fitted to each series by the method of least squares or the method of moments. For the clearing index I assumed the secular trend to be the compound interest law and the function $y = B\ C^t$, where t represents time and B and C are constants determined by the data, was fitted to the series. The deviations of the raw figures from the lines of secular trend were found and designated the "cycles."

TABLE I.

COEFFICIENTS OF CORRELATION FOR THE CYCLES OF THE BUSINESS BAROMETER AND THE CYCLES OF WHOLESALE PRICES, GROSS RECEIPTS OF RAILROADS, CORRECTED CLEARING INDEX AND SURPLUS RESERVES OF NEW YORK BANKS TOGETHER WITH COEFFICIENTS FOR MULTIPLE DIFFERENCES, FIRST TO SIXTH, WITH VARIOUS DEGREES OF LAG, 1879-1913.

Business Barometer Correlated with.	Coefficients of Correlation. (a)						
	r_{-3}.	r_{-2}.	r_{-1}.	r_0.	r_{+1}.	r_{+2}.	r_{+3}.
Wholesale prices, Cycles	....	....	+.90	+.95	+.85	....	....
1st D.	....	....	+.20	+.78	−.03	....	....
2nd D.	....	....	−.22	+.78	−.46	....	....
3rd D.	....	....	−.35	+.78	−.57	....	....
4th D.	....	....	−.40	+.76	−.59	....	....
5th D.	+.12	−.01	−.42	+.74	−.58	+.15	+.09
6th D.	....	....	....	+.71	....	....	....
Gross Receipts of Railroads							
Cycles	....	....	+.91	+.94	+.79	....	....
1st D.	....	....	+.34	+.80	−.28	....	....
2nd D.	....	....	+.08	+.74	−.64	....	....
3rd D.	....	....	−.06	+.74	−.71	....	....
4th D.	....	....	−.18	+.76	−.74	....	....
5th D.	+.25	−.22	−.24	+.75	−.75	+.25	+.09
6th D.	....	....	....	+.74	....	....	....
Corrected Clearing Index							
Cycles	....	....	+.51	+.77	+.80	+.66	....
1st D.	....	....	−.35	+.32	+.52	−.04	....
2nd D.	....	....	−.55	+.23	+.38	−.29	....
3rd D.	....	. .	−.55	+.21	+.27	−.25	....
4th D.	....	....	−.53	+.21	+.18	−.18	....
5th D.	−.36	+.49	−.53	+.24	+.07	−.07	−.20
6th D.	−.48	+.53	−.52	+.28	−.05	+.05	....
Surplus Reserves of New York Banks							
Cycles	....	....	−.27	−.60	−.62	....	....
1st D.	....	....	+.33	−.72	+.02	....	....
2nd D.	....	....	+.56	−.74	+.23	....	....
3rd D.	....	....	+.64	−.75	+.35	....	....
4th D.	....	....	+.71	−.77	+.42	....	....
5th D.	+.28	−.46	+.75	−.78	+.49	−.20	+.17
6th D.	....	....	+.79	−.81	+.59	....	....

(a) The subscript *i* of the coefficient of correlation r indicates the years lag (+) or years previous (−) of the Business Barometer.

Investigation of the 21 series of cycles led to the conclusion that the fluctuations of 9 of them synchronize and hence can logically be combined into a business barometer. Consequently a business barometer is constructed of the 9 series, those numbered 1 to 9 in the list given on the preceding page.* The coefficients of correlation for the cycles of the business barometer and the cycles of wholesale prices, gross receipts of railroads, corrected clearing index and surplus reserves of

* For details of this investigation see the article by the writer in the *American Economic Review*, December, 1916.

New York associated banks, together with the coefficients of correlation for the first to sixth differences, with various degrees of lag for cycles and differences, are given in Table I. Of course there is an element of spurious correlation in the coefficients for the business barometer, wholesale prices, and gross receipts of railroads because the two last named series enter into the barometer. That element is not believed to be large for the cycles and first differences.

Examination of Table I reveals: first, high, positive, and steady coefficients for concurrent items for wholesale prices and gross receipts of railroads, high, negative, and steady coefficients for surplus reserves, low, positive, and fairly steady coefficients for the corrected clearing index; second, all the coefficients for higher differences show a marked tendency to alternate in algebraic sign as successive degrees of lag are taken in either direction.

TABLE II.

PROBABLE ERRORS OF COEFFICIENTS OF CORRELATION FOR SERIES OF 35 AND 46 ITEMS, RESPECTIVELY, AND FOR THEIR MULTIPLE DIFFERENCES, FIRST TO SIXTH.(a)

(Thirty-five Items in Original Series.)

Coefficients of Correlation.	Items.	1st D.	2nd D.	3rd D.	4th D.	5th D.	6th D.
.90	.02	.03	.03	.03	.04	.04	.04
.80	.04	.05	.06	.07	.07	.08	.08
.70	.06	.07	.08	.09	.10	.11	.11
.60	.07	.09	.10	.12	.12	.13	.14
.50	.09	.11	.12	.13	.15	.16	.16
.40	.10	.12	.14	.15	.16	.17	.18
.30	.10	.13	.15	.16	.18	.19	.20
.20	.11	.14	.16	.17	.19	.20	.21
.10	.11	.14	.16	.18	.19	.21	.22
.00	.11	.14	.16	.18	.19	.21	.22

(Forty-six Items in Original Series.)

Coefficients of Correlation.	Items.	1st D.	2nd D.	3rd D.	4th D.	5th D.	6th D.
.90	.02	.02	.03	.03	.03	.03	.04
.80	.04	.04	.05	.06	.06	.06	.07
.70	.05	.06	.07	.08	.09	.09	.10
.60	.06	.08	.09	.10	.11	.11	.12
.50	.08	.09	.11	.12	.13	.13	.14
.40	.08	.10	.12	.13	.14	.15	.16
.30	.09	.11	.13	.14	.15	.16	.17
.20	.10	.12	.14	.15	.16	.17	.18
.10	.10	.12	.14	.15	.17	.18	.19
.00	.10	.12	.14	.16	.17	.18	.19

(a) Computed from formulae developed by O. Anderson and A. Ritchie-Scott. See *Biometrika*, November, 1915, p. 136.

The significance of the various coefficients depends not only on their size but upon the number of items used in the computation. The probable errors for various coefficients based on 35 items in the original series, first to sixth differences are given in Table II.* The probable errors for sixth differences are, approximately, twice those for the original series. A coefficient of, say .45 or more would be significant for the original series and of .65 or more for sixth differences. What is the explanation of the observed steadiness, and of the alternation of sign of coefficients for various degrees of lag? "Student" believes that the steadiness is due to the random distribution, with respect to time, of the differences. The alternation in sign is a phenomenon not noticed, or if noticed not considered, by the writers on the subject.

First, let us consider the phenomenon of alternation in sign. Let one of the original series be $x_o, x_1, x_2 \quad . \quad . \quad . \quad x_{n-1}$, the first differences being $x_1-x_o,\ x_2-x_1, \quad . \quad . \quad . \quad x_{n-1}-x_{n-2}$.† Suppose the first differences alternate in sign at any point so that we have

$$\begin{aligned} x_K-x_{K-1} &= +a \\ x_{K+1}-x_K &= -b \\ x_{K+2}-x_{K+1} &= +c \\ &\text{etc.} \end{aligned}$$

where a, b, c . . . are positive numbers. The second differences are $-b-a$, $+c+b$, $-d-c$, . . ., a series alternating in sign and *larger* numerically than the series from which it is derived. For the portion of the original series, however, where there is no alternation in sign the first differences will be *smaller* numerically than the items from which they are derived. Since the nth differences are derived from the $(n-1)$th differences in the same way that the first differences are derived from the original series we have the following conclusion: If consecutive items of a series alternate in sign the first and higher differences will also alternate in sign and the resulting items will increase numerically as the order of

* These probable errors are computed from theories developed by O. Anderson, *Biometrika*, April, 1914, p. 269. The formulae for probable errors *may* hold for a very large number of items but I doubt their validity for less than 100 items.

† The standard notation for successive finite differences is Δ, Δ^2, Δ^3, etc. That notation is not used in this paper because it would tend to conceal relations brought out in Part III of this paper

the difference increases. A succession of like signs may persist with the first and higher differences but the numbers resulting will be smaller numerically than those resulting where the sign alternates. Where the variate difference method is applied to two short series we may, therefore, expect the terms alternating in size to be of dominating influence upon the coefficient of correlation. Also when a lag is taken in either direction the coefficients will tend to alternate in sign.

TABLE III.

PERCENTAGES FOUND BY TAKING THE RATIO OF THE NUMBER OF CASES IN WHICH SUCCESSIVE ITEMS DIFFER IN SIGN TO THE POSSIBLE NUMBER OF ALTERNATIONS IN SIGN OF SUCCESSIVE ITEMS; VARIOUS SERIES, FIRST TO SIXTH DIFFERENCES, 1879-1913.

Series.	Percentage of Unlike Signs. (a)					
	1st D.	2nd D.	3rd D.	4th D.	5th D.	6th D.
Business barometer	.55	.59	.64	.87	.79	.79
Wholesale prices	.42	.59	.61	.63	.83	.86
Liabilities of commercial failures	.58	.62	.68	.70	.69	.75
Gross receipts of railroads	.33	.66	.68	.70	.76	.82
Percentage of loans to resources of banks	.55	.66	.68	.73	.72	.82
Shares sold on New York Stock Exchange	.58	.62	.71	.70	.79	.82
Immigration	.42	.59	.61	.67	.72	.79
Coal produced	.52	.69	.74	.83	.93	.93
Percentage of cash to deposits in banks	.52	.78	.84	.83	.86	.89
Average price of shares on N. Y. St. Exch.	.55	.75	.74	.77	.76	.86
Corrected Clearing index	.61	.62	.71	.77	.79	.79
Exports of merchandise	.42	.62	.68	.80	.83	.82
Imports of merchandise	.58	.66	.68	.80	.79	.82
Pig-iron produced	.55	.62	.61	.60	.59	.71
Net earnings of railroads	.30	.53	.68	.70	.83	.82
Price of pig-iron	.55	.62	.61	.73	.79	.79
Surplus reserves of New York banks	.45	.66	.64	.70	.69	.71
New railroad mileage constructed	.55	.66	.81	.83	.83	.82
Indices of crop yield, 3-yr. averages	.48	.66	.77	.93	.93	.93
Average of percentages	.50	.64	.70	.75	.79	.82

(a) A + or — item followed by a null item or *vice versa* is counted ½. The possible numbers of unlike signs are 33, 32, 31, 30, 29 and 28 for the differences, 1st to 6th.

Table III shows the tendency of the higher differences of the series here under consideration to alternate in sign. With 34 items in the series of first differences there are 33 possible alternations in sign when each term is compared with the preceding term; there are also 33 possible cases of steadiness in sign. Counting the number of times that the signs of successive terms of each series alternate and expressing the number as a percentage of the possible number (32 for second differences, 31 for third differences, etc.) we have the percentages appearing in the table. For numbers chosen at

TABLE IV.

COEFFICIENTS OF CORRELATION (a) BETWEEN CONCURRENT ITEMS AND THOSE HAVING A LAG OF ONE IN EITHER DIRECTION, FIRST TO FOURTH DIFFERENCE, OF ALL POSSIBLE COMBINATIONS OF THE FOLLOWING SERIES FOR THE PERIOD 1879-1913:

A. BUSINESS BAROMETER.
B. WHOLESALE PRICES.
C. GROSS RECEIPTS OF RAILROADS.
D. CORRECTED CLEARING INDEX.
E. SURPLUS RESERVES OF NEW YORK BANKS.

	AA.			BB.			CC.			DD.			EE.		
Difference.	r_{-1}.	r_0.	r_{+1}.	r_{-1}.	r_0.	r_{+1}.	r_{-1}.	r_0.	r_{+1}.	r_{-1}.	r_0.	r_{+1}.	r_{-1}.	r_0.	r_{+1}.
First	−.13	+1.00	−.13	+.24	+1.00	+.24	+.05	+1.00	+.05	−.31	+1.00	−.31	−.33	+1.00	−.33
Second	−.49	+1.00	−.49	−.43	+1.00	−.43	−.33	+1.00	−.33	−.54	+1.00	−.54	−.54	+1.00	−.54
Third	−.61	+1.00	−.61	−.64	+1.00	−.64	−.44	+1.00	−.44	−.66	+1.00	−.66	−.65	+1.00	−.65
Fourth	−.69	+1.00	−.69	−.71	+1.00	−.71	−.54	+1.00	−.54	−.74	+1.00	−.74	−.71	+1.00	−.71
	AB.			AC.			AD.			AE.			BC.		
First	+.20	+.78	−.03	+.34	+.80	−.28	−.35	+.32	+.52	+.33	−.72	+.02	+.27	+.66	+.07
Second	−.22	+.78	−.46	+.08	+.74	−.64	−.55	+.23	+.38	+.56	−.74	+.23	−.06	+.62	−.40
Third	−.35	+.78	−.57	−.06	+.74	−.71	−.55	+.21	+.27	+.64	−.75	+.35	−.18	+.64	−.52
Fourth	−.40	+.76	−.59	−.18	+.76	−.74	−.53	+.21	+.18	+.71	−.77	+.42	−.26	+.68	−.58
	BD.			BE.			CD.			CE.			DE.		
First	−.10	+.04	+.66	+.19	−.50	−.21	−.34	+.06	+.59	+.41	−.49	−.27	−.27	−.47	+.43
Second	−.17	−.23	+.67	+.45	−.52	+.03	−.42	−.06	+.50	+.59	−.51	−.12	−.11	−.42	+.51
Third	−.14	−.30	+.65	+.51	−.53	+.15	−.36	−.07	+.41	+.62	−.53	.00	−.01	−.46	+.49
Fourth	−.08	−.33	+.64	+.54	−.55	+.23	−.29	−.07	+.32	+.64	−.55	+.08	+.04	−.38	+.46

(a) In this computation the sums of the differences are assumed to be approximately zero. The error resulting from this assumption is small and decreases as the order of difference increases. For the second and higher differences the error resulting is not over .01. For first differences the error will not be over .01 except in those cases where the secular trend differs widely from a straight line, e. g., the compound interest law with a high rate of increase or decrease. In such cases the error may be as much as .04 or .05. However, in these cases' the exact formula was used. The subscript of r indicates that the series designated by the letter first in alphabetical order precedes (−) or lags behind (+) the other series.

random we should expect the first differences to have 50 per cent. steadiness in sign and 50 per cent. alternation, on the average. We get exactly 50 per cent. as the average for the series listed. For second and higher differences the average percentages of alternation are: 64, 70, 75, 79, 82. A marked tendency to alternation in sign is revealed. The cumulative effect of this tendency to alternate in sign as higher differences are taken is also revealed in Table IV. Where the series *A*, *B*, *C*, *D*, and *E* are correlated with themselves (*AA*, *BB*, *CC*, *DD*, *EE*) there is a numerically increasing but negative coefficient for a lag of one item. Where all possible combinations of the five series are taken the coefficients alternate in sign or show a strong tendency to do so with the higher differences.

TABLE V.

COEFFICIENTS OF CORRELATION BETWEEN TWO RANDOM SERIES OF 35 ITEMS EACH AND BETWEEN THEIR MULTIPLE DIFFERENCES, FIRST TO EIGHT, CONCURRENT AND A LAG OF ONE AND OF TWO ITEMS IN EITHER DIRECTION.

Items Paired.	Coefficients of Correlation.				
	r_{-2}.	r_{-1}.	r_0.	r_{+1}.	r_{+2}.
Originals	+.32	+.08	−.11	−.42	−.05
First Differences	+.32	−.03	+.04	−.29	+.01
Second Differences	+.30	−.15	+.12	−.21	+.02
Third Differences	+.28	−.24	+.21	−.19	+.04
Fourth Differences	+.28	−.28	+.23	−.20	+.07
Fifth Differences	+.31	−.32	+.27	−.18	+.10
Sixth Differences	+.31	−.34	+.29	−.21	+.13
Seventh Differences	+.31	−.36	+.32	−.26	+.18
Eighth Differences	+.30	−.36	+.34	−.30	+.22

This theory of the tendency of signs of terms of higher differences to alternate and, therefore, to affect the coefficients of correlation was tested by applying the method of variate differences to two random series of 35 items each. The method of selection of the numbers was this: the pages of a table of six-place logarithms were turned at random, the tip of a pointer was placed at random on the page and the two digits at the right of the logarithm indicated by the pointer were taken as the items of the series. The coefficients of correlation between the two random series of 35 items and between their multiple differences, first to eighth, concurrent

and for one and two items lag in each direction are given in Table V. For the first and higher differences there is a persistent alternation in sign as we take a lag in either direction.* The coefficients alternate in sign, of course, because, first, the two series correlated alternate in sign, second, the terms alternating in sign become the dominating ones when the products of corresponding items are taken and third, lagging either series in either direction will bring a different set of signs into correspondence. The term Σxy tends to alternate in sign and hence r does.

III.

Does this phenomenon of the alternation in sign of the coefficients, left to right in the tables, have any bearing on the steadiness or unsteadiness of the coefficients, in the tabular columns, based on successive differences but with the same lag throughout? This question will now be considered.

The series of nth differences is derived from the series of $(n-1)$th differences by the same process that the series of first differences is derived from the original figures. Therefore, the expression for the coefficient of correlation for nth difference is the same function of the $(n-1)$th differences that the first differences is of the original series.

Let $r_L^{(m)}$ represent the coefficient of correlation between the mth differences of the series $x_o, x_1, \quad . \ . \ . \quad x_{n-1}$ and $y_o, y_1, \ . \ . \ . \ y_{n-1}$ where the subscript L denotes the lag of the x series. When $L=1$ we have the pairs $x_1 y_o$, $x_2 y_1 \ . \ . \ . \ x_{n-1}\, y_{n-2}$; when $L=-1$ we have the pairs $x_o y_1$, $x_1 y_2 \ . \ . \ . \ x_{n-2}\, y_{n-1}$, etc. The formula for the coefficient of correlation between concurrent items of the original series is usually written in the form

$$r_o = \frac{\sum\limits_{K=o}^{K=n-1} (x_K - \overline{x})(y_K - \overline{y})}{\sqrt{\sum\limits_{K=o}^{K=n-1} (x_K - \overline{x})^2 \sum\limits_{K=o}^{K=n-1} (y_K - \overline{y})^2}} \tag{1}$$

when $\overline{x}$ and $\overline{y}$ are arithmetic averages of the respective series.

* None of the coefficients found signify appreciable correlation between the series. In but one case is the coefficient more than three times its probable error and in the majority of cases the probable error is approximately the same as the coefficient.

The function r_o may be written

$$r_o=\frac{n\sum\limits_{o}^{n-1}x_Ky_K-\sum\limits_{o}^{n-1}x_K\sum\limits_{o}^{n-1}y_K}{\sqrt{[n\sum\limits_{o}^{n-1}x_K^2-(\sum\limits_{o}^{n-1}x_K)^2][n\sum\limits_{o}^{n-1}y_K^2-(\sum\limits_{o}^{n-1}y_K)^2]}} \tag{2}$$

The coefficient of correlation r'_o between concurrent first differences x_1-x_o, y_1-y_o; x_2-x_1, y_2-y_1; . . . $x_{n-1}-x_{n-2}$, $y_{n-1}\ y_{n-2}$ is, noticing that $\sum\limits_{1}^{n-1}(x_K-x_{K-1})=x_{n-1}-x_o$ and $\sum\limits_{1}^{n-1}(y_K-y_{K-1})=y_{n-1}-y_o$, *in terms of the original items.*

$$r'_o=\frac{(n-1)\sum\limits_{1}^{n-1}(x_K-x_{K-1})(y_K-y_{K-1})-(x_{n-1}-x_o)(y_{n-1}-y_o)}{\left\{\sqrt{[(n-1)\sum\limits_{1}^{n-1}(x_K-x_{K-1})^2-(x_{n-1}-x_o)^2]}\cdot\sqrt{[(n-1)\sum\limits_{1}^{n-1}(y_K-y_{K-1})^2-(y_{n-1}-y_o)^2]}\right\}} \tag{3}$$

Assuming $x_{n-1}-x_o$ and $y_{n-1}-y_o$ to be negligible in comparison with other terms of the function (called assumption a) we will discard the former. The function becomes

$$r'_o=\frac{2\sum\limits_{o}^{n-1}x_Ky_K-(x_oy_o+x_{n-1}y_{n-1})-(\sum\limits_{1}^{n-1}x_{K-1}y_K+\sum\limits_{1}^{n-1}x_Ky_{K-1})}{\left\{\sqrt{[2\sum\limits_{o}^{n-1}x_K^2-(x_o^2+x^2_{n-1})-2\sum\limits_{1}^{n-1}x_Kx_{K-1}]}\cdot\sqrt{[2\sum\limits_{o}^{n-1}y_K-(y_o^2+y_{n-1}^2)-2\sum\limits_{1}^{n-1}y_Ky_{K-1}]}\right\}} \tag{4}$$

Assuming that the original items of each series are random in order, first with respect to the adjacent items of the same series and, second, with respect to the items adjacent to the concurrent item of the other series, that is assuming $\sum\limits_{1}^{n-1}x_{K-1}y_K$, $\sum\limits_{1}^{n-1}x_Ky_{K-1}$, $\sum\limits_{1}^{n-1}x_Kx_{K-1}$, and $\sum\limits_{1}^{n-1}y_Ky_{K-1}$ all equal zero (called assumption b) we have

$$r'_o=\frac{2\sum\limits_{o}^{n-1}x_Ky_K-(x_oy_o+x_{n-1}y_{n-1})}{\sqrt{[2\sum\limits_{o}^{n-1}x_K^2-(x_o^2+x_{n-1}^2)][2\sum\limits_{o}^{n-1}y_K^2-(y_o^2+y_{n-1}^2)]}} \tag{5}$$

Assuming $x_o y_o + x_{n-1} y_{n-1}$, $x_o^2 + x_{n-1}^2$ and $y_o^2 + y_{n-1}^2$ to be negligible in comparison with other terms in the function (called assumption *c*) we have, discarding the three terms named,

$$r'_o = \frac{\sum_o^{n-1} x_K y_K}{\sqrt{\sum_o^{n-1} x_K^2 \cdot \sum_o^{n-1} y_K^2}} = r_o, \text{ approximately, if } \bar{x} = \bar{y} = o$$

That is, if at any time assumptions *a*, *b*, and *c* hold true for any series of multiple differences, the coefficient of correlation for their first differences will equal the coefficient for the items from which the first differences were derived. If the same assumptions (*a*, *b*, and *c*) hold true for the series of first differences then $r''_o = r'_o$. Further if the assumptions hold true repeatedly for successive differences from and after the *p*th, we have $(m > p)$

$$r^{(m)} = r^{(m-1)} = \quad . \quad . \quad . \quad = r^{(p+1)} = r^{(p)}$$

It is obvious then, that the coefficients of correlation for successive differences will remain stable if assumptions *a*, *b*, and *c* hold true for successive differences. The condition (assumption that *a*, *b*, and *c* hold true repeatedly) is sufficient to produce stability, but is it necessary?

Consider form (4) of the function r'_o. If the two terms, $\sum_1^{n-1} x_K x_{K-1}$ and $\sum_1^{n-1} y_K y_{K-1}$, appearing in the denominator are negative in sign, if the expression $(\sum_1^{n-1} x_{K-1} y_K + \sum_1^{n-1} x_K y_{K-1})$ is opposite in sign to $\sum_1^{n-1} x_K y_K$, and if the terms are of appropriate size as well as sign, if these assumptions hold true, dropping $\sum_1^{n-1} x_K x_{K-1}$, $\sum_1^{n-1} y_K y_{K-1}$, and $(\sum_1^{n-1} x_{K-1} y_K + \sum_1^{n-1} x_K y_{K-1})$ will not affect the value of r'_o. Also r'_o will approximately equal r_o. In other words the fulfilling of the assumptions named will result in stable coefficients for successive differences. It is my contention that for a moderate number of items ($n = 35$ or 40) the conditions here specified are apt to occur and be the cause of any stability of the coefficients of correlation between multiple differences.

TABLE VI.

COEFFICIENTS OF CORRELATION BETWEEN THE BUSINESS BAROMETER AND THE CLEARING INDEX FOR THE UNITED STATES, FIRST TO SIXTH DIFFERENCES, CONCURRENT AND LAG AT ONE AND TWO ITEMS IN EITHER DIRECTION, 1879-1913.

Difference.	Coefficients of Correlation.			
	r_{-1}.	r_0.	r_{+1}.	r_{+2}.
First	−.33	+.68	+.39	−.22
Second	−.65	+.59	+.17	−.35
Third	−.71	+.57	+.02	−.26
Fourth	−.72	+.57	−.07	−.19
Fifth	−.71	+.57	−.14	−.06
Sixth	−.71	+.58	−.25	+.11

The results of applying the variate difference correlation method to the business barometer and the clearing index (not corrected for prices) are tabulated in Table VI. The first to sixth differences were used and the items were lagged in both directions. The stability of the coefficients for concurrent items second to sixth differences (r^{ii}_o, r^{iii}_o, r^{iv}_o, r^{v}_o, r^{vi}_o) and the instability of the coefficients for one item lag in the business barometer (r^{iii}_{+1} to r^{vi}_{+1}) are noticeable. Let us investigate the stability of r^{ii}_o and r^{iv}_o and the instability of r^{iii}_{+1} and r^{v}_{+1}. Using from (4) we have these values:

Term.		r^{iii}_o	r^{iv}_o	r^{iii}_{+1}	r^{v}_{+1}
$2\Sigma x_K y_K$	=	+3134	+2718	+ 888	− 950
$(x_o y_o + x_{n-1} y_{n-1})$	=	− 10	− 10	− 9	+ 12
$(\Sigma x_{K-1} y_K + \Sigma x_K y_{K-1})$	=	−1293	− 1618	+ 681	+ 2506
$2\Sigma x^2_K$	=	+7918	+ 2126	+7676	+ 6128
$(x^2_o + x^2_{n-1})$	=	+ 122	+ 58	+ 122	+ 37
$2\Sigma x_K x_{K-1}$	=	−3132	− 1276	−2890	− 4254
$2\Sigma y^2_K$	=	+3596	+10536	+3498	+32814
$(y^2_o + y^2_{n-1})$	=	+ 50	+ 250	5	+ 1360
$2\Sigma y_K y_{K-1}$	=	−1722	− 6298	−1694	−21556

Computing r_o^{iii}, r_o^{iv} and r_{+1}^{iii} and r_{+1}^{v} in two ways, by form (4) and form (5), we have the following results:

	Form (4).	Form (5).
r_o^{iii}	+.59	+.59
r_o^{iv}	+.58	+.59
r_{+1}^{iii}	+.03	+.17
r_{+1}^{v}	−.15	−.07

The stability of r_o^{iii} and r_o^{iv} is explainable by the balancing effect of items $(\Sigma x_{K-1}y_K+\Sigma x_K y_{K-1})$, $\Sigma x_K x_{K-1}$ and $\Sigma y_K y_{K-1}$, the instability of r_{+1}^{iii} and r_{+1}^{v} is explainable by the lack of balance between those items as they appear in the function. The items named are not approximately zero, in the case of r_o^{iii} and r_o^{iv}, as "Student" assumes, but numerically as large as $\Sigma x_K y_K$.

TABLE VII.

COEFFICIENTS OF CORRELATION BETWEEN CONCURRENT ITEMS OF BUSINESS BAROMETER AND PRICES, RAILROAD GROSS EARNINGS, CORRECTED CLEARING INDEX AND SURPLUS RESERVES OF NEW YORK BANKS COMPUTED (I) BY EXACT FORMULA, (II) BY DISCARDING $\Sigma(x_K-x_{K-1})$ AND $\Sigma(y_K-y_{K-1})$ AND (III) BY DISCARDING $\Sigma x_{K-1}y_K-\Sigma x_K y_{K-1}$; $\Sigma x_K x_{K-1}$ AND $\Sigma y_K y_{K-1}$ FROM FORMULA.

	Prices.			R. R. Gross Ear.			Corr. Cl. Index.			Surplus Reserves.		
	I	II(a)	III(b)	I	II(a)	III(b)	I	II(a)	III(b)	I	II(a)	III(b)
Cycles	+.95	+.94	..	+.94	+.94	..	+.77	+.76	..	−.60	−.60	..
1st D.	+.78	+.78	+.94	+.80	+.80	+.94	+.32	+.32	+.76	−.72	−.71	−.60
2nd D.	+.78	+.78	+.79	+.74	+.74	+.81	+.23	+.23	+.34	−.74	−.74	−.72
3rd D.	+.78	+.78	+.79	+.74	+.74	+.77	+.21	+.21	+.23	−.75	−.75	−.74
4th D.	+.76	+.76	+.78	+.76	+.76	+.76	+.21	+.21	+.22	−.77	−.77	−.75
5th D.	+.74	+.74	+.76	+.75	+.75	+.76	+.24	+.24	+.22	−.78	−.78	−.77
6th D.	+.71	+.71	+.73	+.74	+.74	+.74	+.28	+.28	+.26	−.81	−.81	−.79

(a) The results in columns II coincide with those in columns I because the sums of the variate differences are approximately zero. See Table VIII.

(b) The results in columns III are equal to those of columns II of a lower order of difference. Approximate steadiness is reached not because the terms discarded are approximately zero but because they "balance." Evidence that the terms discarded are large is given by the coefficients of correlation for a lag in Table IV.

Table VII gives the results of applying, (I) the exact formula, (II) the formula obtained by assuming $\Sigma(x_K - x_{K-1}) = 0$, $\Sigma(y_K - y_{K-1}) = 0$ and (III) the formula obtained by assuming that the terms $(\Sigma x_{K-1}y_K - \Sigma x_K y_{K-1})$, $\Sigma x_K x_{K-1}$ and $\Sigma y_K y_{K-1}$ balance and therefore may be discarded, of applying these formulae to the business barometer and each of the series, wholesale prices, railroad gross earnings, corrected clearing index, and surplus reserves of New York associated banks. Form II gives coefficients practically identical with those resulting from the exact Form I showing that the assumption involved can safely be made. Table VIII illustrates the negligibility of the algebraic sums by comparing the algebraic sums of the second differences of various series with the absolute sums of the same items. Form III of Table VII gives coefficients which are equal to those for preceding differences as would be expected.

TABLE VIII.

ALGEBRAIC SUMS AND ABSOLUTE SUMS OF THE SECOND DIFFERENCES OF SIX SERIES WITH RATIO OF FORMER TO LATTER.

Second Differences of:	Algebraic Sum.	Absolute Sum.	Ratio.
Business barometer	0	150	.00
Wholesale prices	−11	157	−.07
Railroad gross earnings	+22	312	+.07
Clearing index	−7	277	−.03
Corrected clearing index	−6	456	−.01
Surplus reserves of New York banks	+2	568	.00

Consideration of the effects that assumptions *a*, *b*, and *c* have upon the value of the coefficients for multiple differences, leads to the conclusion that assumptions *a* and *c* are in accordance with the facts and that making them will not, for series of 35 or more items, in general affect the coefficient of correlation by more than .01. Assumption *b*, however, *does not* usually hold true. Nevertheless, stability is frequently secured by a balance between the terms appearing in numerator and denominator. Instability of the coefficients is explainable by lack of such balance. The fulfilling of assumption *c* is sufficient but not necessary to secure stability. Moreover, stability is usually secured when assumption *c* does not hold

true. In view of this conclusion, what significance has the variate difference correlation method for short economic time series? This question will now be considered.

IV.

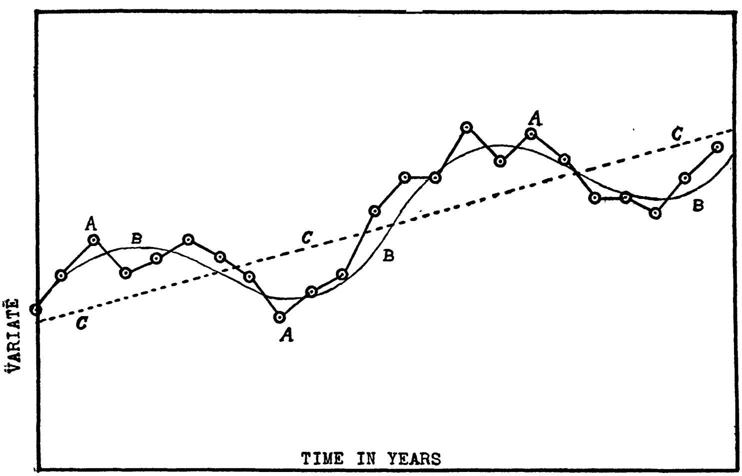
The items of annual time series of economic data may be conceived to be constituted of the following elements or component parts:

First, the *secular trend* or growth element due to the increase of population and development of industry;

Second, *cyclical fluctuations*, extending over a number of years and having a greater or less degree of periodicity, due to the alternating periods of business prosperity and depression;

Third, *irregular fluctuations* from year to year due to the influence of accidental or, at any rate, unpredictable events such as inventions, striking changes in fashion, or war.

FIGURE 1



A—REPRESENTS ECONOMIC DATA.
B—REPRESENTS CYCLICAL FLUCTUATION.
C—REPRESENTS SECULAR TREND.

An idealistic representation of a time series of economic data containing the three elements, secular trend, cyclical, and irregular fluctuation is given in Figure 1. For particular data the secular trend may, of course, be other than a straight line and the cyclical fluctuations other than the simple sine curve.

The concept of secular trend or normal growth element for which I contend is that of an element which increases or decreases regularly according to some principle. That principle may be linear, *i. e.*, the addition of a constant amount in each time interval as assumed in Figure 1, or it may be the compound interest law, the addition of an equal percentage in each time interval, or it may be a second degree parabola, or some other law. In any case, however, the mathematical function assumed to represent the secular trend should not "fit" the cyclical or irregular fluctuations of the data. Now the process of taking multiple differences is equivalent, on the assumption that the series is an algebraic function of time, to reducing the degree of the function of t by one for each difference taken. Thus, if the secular trend be linear, if a straight line should be fitted to the data and if the deviations of the original series from the corresponding ordinates of the straight line are taken (the cycles), then the coefficient of correlation between the first differences of the original items and that between the first differences of the cycles are identical. This theorem may be proven as follows:
Let $x_o, x_1, x_2 \quad . \quad . \quad . \quad x_{n-1}$ be the original series of n items and let $x=mt+b$ be the line of secular trend. The first differences are of the form x_K-x_{K-1}. The cycles are of the form $x_K-(mt_K+b)$. The first differences of the cycles are of the form $x_K-x_{K-1}-m(t_K-t_{K-1})$. Since t_K-t_{K-1} is a constant for all values of K (being the time interval for which items are taken) the first differences of the cycles differ from the first differences of the original series by a constant. Hence the coefficients of correlation obtained from them are identical.

The coefficient of correlation between second differences is identical with that found between the corresponding deviations of items from the smoothed curve obtained by taking

three-year averages.* This theorem may be proven as follows:

Let x_o, x_1, x_2 . . . x_{n-1} be the original series. The second differences are of the form $x_{K+1}-2x_K+x_{K-1}$. The deviation of any items from the three-year average is $x_K-\frac{x_{K+1}+x_K+x_{K-1}}{3}$ or $\frac{-x_{K+1}+2x_K-x_{K-1}}{3}$. The form just given is obtainable from the form for second differences by multiplying by $-\frac{1}{3}$. Therefore the coefficients of correlation between second differences are the same as those between deviations from three-year averages.†

* Moore has correlated deviations from three-year averages of crop yield per acre (weighted index of nine leading crops) and production of pig-iron. He does not appear to realize that his coefficients are the same as those for second differences. The coefficient for crops and pig-iron production with a one-year lag of the latter (r''_{+1}) for the period 1870–1911 is .254; for crops and general prices for the same period the coefficient (r''_0) is .303. See Moore's *Economic Cycles*, pp. 107, 118.

† (Theorem for form of mth difference)

Theorem: The mth difference of the terms of a series

$$x_1, x_2, x_3 \quad . \quad . \quad . \quad x_n$$

is of the form

$$x_K^{(m)}=[C_m x_{K+m}-C_{m-1}x_{K+m-1}+C_{m-2}x_{K+m-2}- \quad . \quad . \quad . \quad +(-1)^m C_o x_K]$$

where C_i indicates the number of combinations of m things taken i at a time.

The first differences of the series

x_1, x_2 . . . x_n are x_2-x_1, x_3-x_2 . . . $x_{K+1}-x_K$ . . . x_n-x_{n-1}.

The second differences are

$$x_3-2x_2+x_1 \quad . \quad . \quad . \quad x_{K+1}-2x_K+x_{K-1} \quad . \quad . \quad . \quad x_n-2x_{n-1}+x_{n-2}.$$

Assume the mth difference to be of the form

$$x_{K+m}-mx_{K+m-1}+\frac{m(m-1)}{\underline{|2}}x_{K+m-2}- \quad . \quad . \quad . \quad +(-1)^m x_K \text{ and}$$

$$x_{K+m+1}-mx_{K+m}+\frac{m(m-1)}{\underline{|2}}x_{K+m-1}- \quad . \quad . \quad . \quad +(-1)_m x_{K+1}$$

Taking the difference we have

$$x_{K+m+1}-(m+1)x_{K+m}+\frac{m(m-1)-2m}{\underline{|2}}x_{K+m-1} \quad . \quad . \quad .$$

$$. \quad . \quad . \quad \pm\left[\frac{\underline{|m}}{\underline{|i}\ \underline{|m-i}}+\frac{\underline{|m}}{\underline{|i-1}\ \underline{|m-i+1}}\right]x_{K+m-i+1} \quad . \quad . \quad .$$

Consider the coefficient of the $(i+1)$th term:

$$\frac{\underline{|m}}{\underline{|i}\ \underline{|m-i}}+\frac{\underline{|m}}{\underline{|i-1}\ \underline{|m-i+1}}, \text{which reduces to}$$

$$\underline{|m}\left[\frac{(m-i+1)+i}{\underline{|i}\ \underline{|m-i+1}}\right]=\frac{\underline{|m+1}}{\underline{|i}\ \underline{|m+1-i}}=C_i^{m+1}$$

Since we have shown that the 2nd differences are expressions with the binomial coefficients corresponding to the order of difference and since, *assuming* that the mth difference appears with the binomial coeffi-

These theoretical considerations taken in connection with applications and tests of the variate difference method lead me to the following conclusions concerning the meaning of the coefficients of correlation for the raw figures and their multiple differences:

In general, significant coefficients of correlation for the raw figures of two time series indicate the similarity of the growth elements of the two series, *if large growth elements exist.* The existence or non-existence of such elements is readily determined graphically or by fitting a simple function to the data.

Significant coefficients of correlation for first differences indicate that the cyclical fluctuations synchronize, *if there be cyclical fluctuations.* Evidence of such cycles may be secured by plotting the deviations from the assumed secular trend.

Significant coefficients of correlation for second and in some cases higher differences indicate, in general, that the irregular fluctuations synchronize. Coefficients for higher differences of short series contain a large spurious element which increases with the order of difference. This element is due to the tendency of the items to alternate in sign.

These conclusions assume that the magnitude of the elements due to secular trend, business cycles, and irregular events vary in the order in which these elements are named. In case we are dealing with data having marked cycles (of the same variety) and are interested in the correlation of the cycles the coefficients for the cycles and first differences constitute the proper basis for judgment rather than coefficients for higher differences. Stability of coefficients for higher differences, in such a case, probably means that the influence of all the fluctuations except the irregular ones, *i. e.*, the oscillations, has been eliminated by the variate difference process.

cients corresponding to the terms of the expansion of $(1+1)^m$, we have proven that the $(m+1)$th difference has as coefficients the terms of the expansion of $(1+1)^{m+1}$ then, by mathematical induction, the proposition stated is true.

Professor A. A. Young has called my attention to the fact that the foregoing theorem is demonstrated in the *Institute of Actuaries' Text Book*, 2nd ed., Part II, p. 427.

TABLE IX.

LONDON CLEARINGS (COLUMN 1) WITH THE 9-YEAR MOVING AVERAGE (2) AND DEVIATIONS FROM MOVING AVERAGE (3) STRAIGHT LINE (4), PARABOLA (5) AND COMPOUND INTEREST LAW (6); SAUERBECK'S PRICES (7) WITH THE 9-YEAR MOVING AVERAGE (8) AND DEVIATIONS FROM MOVING AVERAGE (9), STRAIGHT LINE (10) AND PARABOLA (11); 1868-1913. ALSO DEVIATIONS OF LONDON CLEARINGS FROM THE TWO STRAIGHT LINES FITTED TO DATA FOR 1868-1896 AND FOR 1897-1913 (12) AND DEVIATIONS OF SAUERBECK'S PRICES FROM TWO STRAIGHT LINES SIMILARLY FOUND (13).

Year.	1.(a)	2.	3.	4.	5.	6.	7.(b)	8.	9.	10.	11.	12.	13.
1868	34	....	....	+6	−18	−3	99	....	....	+8	−11	−10	−7
1869	36	....	....	+5	−15	−2	98	....	....	+8	−10	−9	−6
1870	39	....	....	+6	−11	−1	96	....	....	+6	−9	−7	−6
1871	48	....	....	+13	−2	+7	100	....	....	+11	−2	+1	−1
1872	59	49	+10	+21	+10	+17	109	101	+8	+21	+10	+11	+10
1873	61	51	+10	+21	+12	+17	111	100	+11	+23	+14	+12	+13
1874	59	53	+6	+17	+10	+15	102	99	+3	+15	+7	+9	+6
1875	57	54	+3	+12	+9	+11	96	97	−1	+10	+4	+6	+2
1876	50	55	−5	+3	+1	+2	95	96	−1	+9	+5	−2	+2
1877	50	55	−5	+1	+1	+1	94	93	+1	+9	+6	−3	+3
1878	50	56	−6	−1	+1	−1	87	90	−3	+3	+1	−4	−2
1879	49	56	−7	−5	−1	−3	83	88	−5	−1	−2	−7	−5
1880	58	56	+2	+2	+8	+4	88	86	+2	+5	+5	+1	+2
1881	64	56	+8	+6	+14	+8	85	83	+2	+2	+4	+6	+1
1882	62	57	+5	+1	+10	+5	84	81	+3	+2	+4	+3	+1
1883	59	58	+1	−4	+6	0	82	79	+3	+1	+4	−1	+1
1884	58	61	−3	−7	+3	−3	76	77	−1	−5	−1	−3	−3
1885	55	62	−7	−13	−1	−8	72	75	−3	−8	−4	−7	−6
1886	59	64	−5	−11	+2	−6	69	74	−5	−10	−5	−4	−7
1887	61	65	−4	−11	+2	−6	68	73	−5	−11	−5	−3	−7
1888	69	65	+4	−6	+8	+1	70	71	−1	−8	−2	+4	−3
1889	76	66	+10	−1	+13	+5	72	70	+2	−6	+1	+10	+1
1890	78	67	+11	−1	+13	+5	72	69	+3	−5	+1	+10	+2
1891	69	69	0	−13	+2	−7	72	68	+4	−4	+2	0	+4
1892	65	71	−6	−19	−5	−13	68	68	0	−8	−2	−5	+2
1893	65	71	−6	−21	−7	−15	68	67	+1	−7	−1	−6	+3
1894	63	72	−9	−25	−12	−19	63	66	−3	−11	−6	−9	0
1895	76	73	+3	−15	−2	−9	62	65	−3	−12	−7	+3	+1
1896	76	76	0	−17	−5	−12	61	66	−5	−12	−8	+2	+1
1897	75	79	−4	−20	−9	−16	62	66	−4	−10	−7	0	−2
1898	81	83	−2	−17	−7	−12	64	66	−2	−8	−5	+1	−2
1899	92	87	+5	−8	+1	−5	68	67	+1	−3	−1	+7	+1
1900	90	91	−1	−12	−5	−10	75	68	+7	+4	+6	−1	+7
1901	96	96	0	−9	−2	−7	70	69	+1	0	+1	0	+1
1902	100	102	−2	−7	−3	−5	69	70	−1	0	−1	−1	−1
1903	101	107	−6	−8	−6	−8	69	72	−3	0	−1	−6	−2
1904	106	110	−4	−6	−5	−7	70	73	−3	+2	−1	−6	−2
1905	123	115	+8	+9	+8	+7	72	73	−1	+5	0	+6	−2
1906	127	121	+6	+11	+7	+8	77	74	+3	+10	+4	+5	+2
1907	127	126	+1	+9	+2	+4	80	75	+5	+14	+6	−1	+4
1908	121	132	−11	0	−8	−6	73	76	−3	+7	−2	−12	−4
1909	135	139	−4	+12	+1	+5	74	78	−4	+9	−2	−3	−4
1910	147	....	....	+22	+8	+12	78	....	....	+14	+1	+3	−1
1911	146	....	....	+18	+1	+7	80	....	....	+16	+1	−3	−1
1912	160	....	....	+30	+10	+16	85	....	....	+22	+5	+6	+3
1913	164	....	....	+32	+8	+17	85	....	....	+23	+3	+4	+2

(a) In £100,000,000. (b) Relative Indices.

Judgment concerning the correlation of cyclical fluctuations of two series must be preceded by elimination of the secular trend. The choice of a function to represent the secular trend, indeed the choice of the method of eliminating the trend,

whether by curve fitting or otherwise, these are questions fundamental to the process. I will test the effect that various suppositions concerning the secular trend have upon the correlation coefficients between the deviations from the various trends (cycles) resulting from the suppositions. The two series chosen for this test are London clearings and Sauerbeck's price indices given with other data in Table IX. These series are chosen, first, because the secular trends are dissimilar, second, because the trends differ most widely from the linear of any which could be found, and, third, because the variate difference correlation coefficients for these series are puzzling to the author of the variate difference correlation method. "Student" applied his method to London clearings per capita and Sauerbeck's prices and to marriage rate and wages, finding the following coefficients:

I. CLEARINGS AND PRICES.

Raw Figures.	1st D.	2nd D.	3rd D.	4th D.	5th D.	6th D.
−.33	+.51	+.30	+.07	+.11	+.05	..

II. MARRIAGE RATE AND WAGES.

−.52	+.67	+.58	+.52	+.55	+.58	+ 55

He says, "The difference between I and II is very marked, and would seem to indicate that the causal connection between index numbers and Bankers' clearing house rates is not altogether of the same kind as that between marriage rate and wages, though all four variables are commonly taken as indications of the short period trade wave."*

Figure 2 presents Sauerbeck's indices with linear and parabolic secular trends, the functions being fitted to the data by the method of moments. Figure 3 presents London bank clearings with linear, parabolic, and exponetial functions fitted to the data. Figure 4 presents both series with their respective nine-year moving averages, nine years being deter-

* *Biometrika*, April, 1914, p. 180.

mined by inspection as the length of the business wave or cycle. Figures 5, 6, 7, and 8 present the deviations, positive or negative, of the two series from the various secular trends. The last named figures all show a striking correspondence of the cyclical fluctuations of the two series. It will be noticed that fluctuations of clearings show a tendency to precede or forecast the fluctuations in prices.

Figures 5, 6, 7, and 8 throw some light on Babson's hypothesis that economic action and reaction are equal, *i. e.*, that consecutive areas above and below the line of normal growth should be equal for a correct normal line. It is true that the *sum* of areas above and below the lines are roughly equal; that, the method of curve fitting accomplishes. But what are *consecutive* areas, the long-time areas of Figures 6 and 8 or the short-time areas of Figures 5 and 7? It is obvious that we would get still more heterogeneous results, as regards positive and negative areas, if we should use series of various lengths, say 20 years or 70 years, or series of monthly or quarterly rather than annual data.

FIGURE 2.

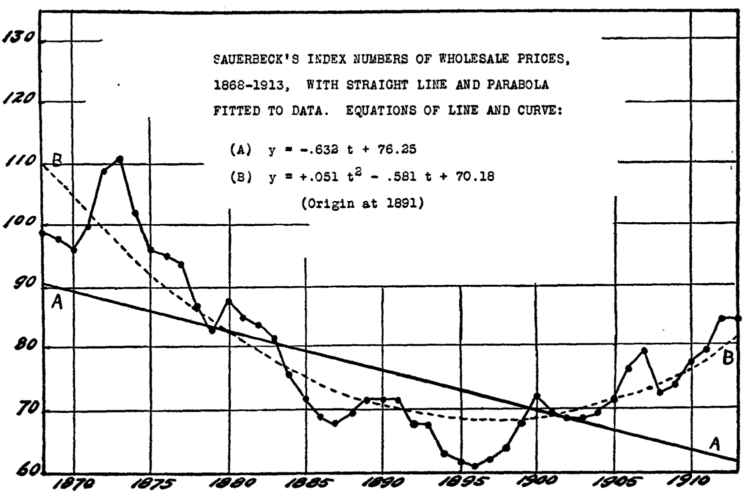


FIGURE 3

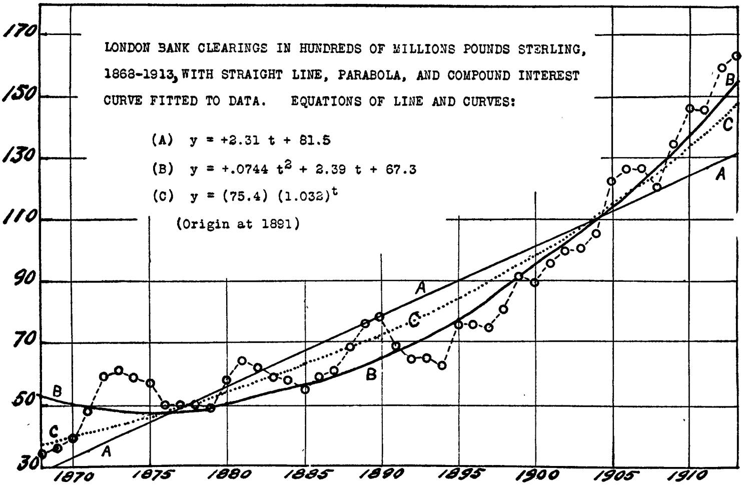
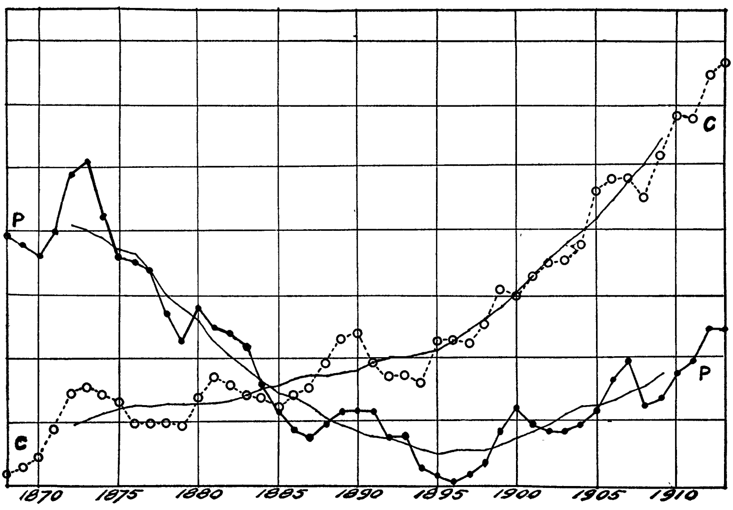
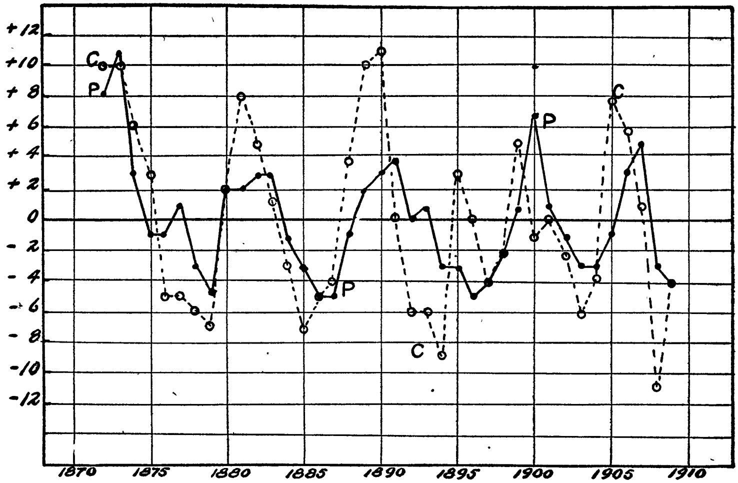


FIGURE 4



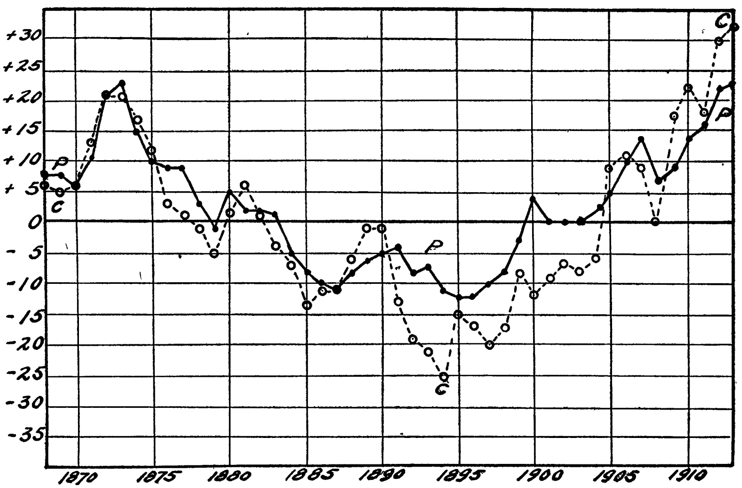
SAUERBECK'S PRICE INDICES (P) AND LONDON CLEARINGS (C), 1868-1913, WITH THEIR RESPECTIVE NINE-YEAR MOVING AVERAGES, 1872-1909.

FIGURE 5



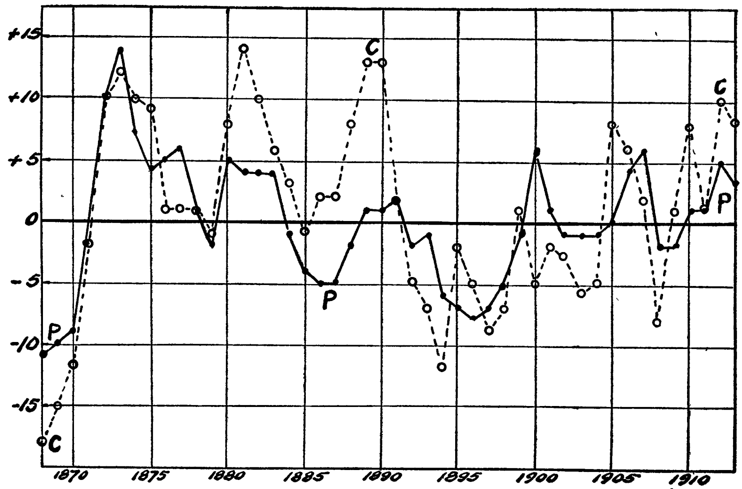
DEVIATIONS OF LONDON CLEARINGS (C) AND SAUERBECK'S PRICE INDICES (P) FROM THEIR RESPECTIVE NINE-YEAR MOVING AVERAGES AS SECULAR TRENDS, 1872–1909.

FIGURE 6



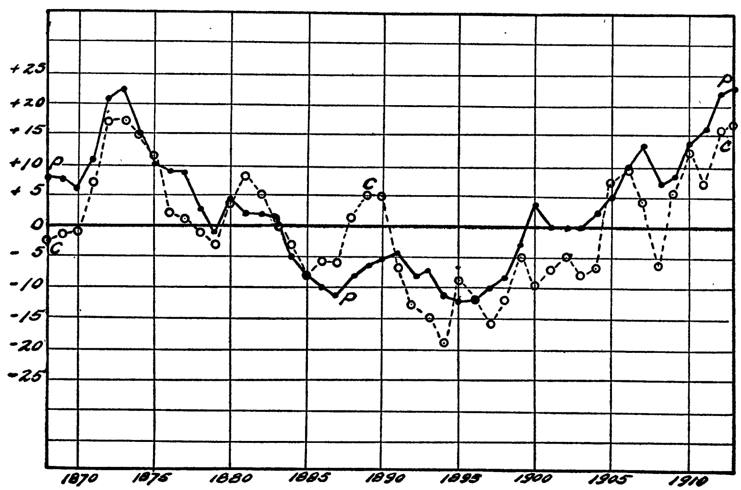
DEVIATIONS OF LONDON CLEARINGS (C) AND SAUERBECK'S PRICES (P) FROM THEIR RESPECTIVE LINEAR SECULAR TRENDS, 1868–1913.

FIGURE 7



DEVIATIONS OF LONDON CLEARINGS (C) AND SAUERBECK'S PRICES (P) FROM THEIR RESPECTIVE PARABOLIC SECULAR TRENDS, 1868–1913.

FIGURE 8



DEVIATIONS OF LONDON CLEARINGS FROM TREND AS COMPOUND INTEREST CURVE (C) AND AT SAUERBECK'S PRICES FROM LINEAR TREND (P), 1868–1913.

TABLE X.

COEFFICIENTS OF CORRELATION BETWEEN SAUERBECK'S PRICE INDICES AND LONDON CLEARINGS, 1868-1913.

A. Raw Figures and their Differences.
B. Deviations from 9-year Moving Average and Differences.
C. Deviations from Straight Lines and Differences.
D. Deviations from Parabolas and Differences.
E. Deviations from Compound Interest Law for Clearings and Straight Line for Prices and Differences.

	r_{-2}.	r_{-1}.	r_0.	r_{+1}.	r_{+2}.
A.					
Raw Figures	....	−.37	−.31	−.28	....
First Differences	....	+.05	+.42	+.37	+.12
Second Differences	....	−.20	+.31	+.14	....
Third Differences	....	−.27	+.26	+.03	....
Fourth Differences	....	−.26	+.23	−.01	....
Fifth Differences	....	−.23	+.23	−.06	....
Sixth Differences	....	−.21	+.21	−.08	....
B.					
Deviations	−.27	+.19	+.64	+.69	+.36
First Differences	....	−.06	+.42	+.37	....
Second Differences	....	−.15	+.22	+.10	....
Third Differences	....	+.10	+.15	+.02	....
C.					
Deviations	+.75	+.85	+.92	+.90	+.80
First Differences	....	+.09	+.55	+.43	....
Second Differences	....	−.21	+.35	+.08	....
Third Differences	....	−.26	+.31	−.04	....
D.					
Deviations	+.20	+.49	+.73	+.65	+.33
First Differences	....	+.06	+.49	+.36	....
Second Differences	....	−.22	+.35	+.06	....
Third Differences	....	−.28	+.31	−.06	....
E.					
Deviations	+.67	+.75	+.82	+.76	+.59
First Differences	....	+.02	+.50	+.38	....
Second Differences	....	−.26	+.33	+.13	....
Third Differences	....	−.32	+.35	.00	....

The main question upon which we wish to get light is, however, the effect of the various methods of eliminating the secular trend upon the coefficients of correlation between corresponding deviations. Table X gives the coefficients of

correlation between Sauerbeck's indices and London clearings taking the raw figures and their differences, first to sixth, and also taking deviations from various secular trends and their differences, first to third. Coefficients are presented for concurrent items and for a lag in both directions, a lag of one for differences and a lag of one and of two for deviations or cycles.

The coefficients of correlation for the raw figures (−.37, −.31 and −.28) show that the secular trends of prices and clearings are in opposite directions. The coefficients for the first differences of the raw figures and of all the deviations indicate an appreciable positive correlation for concurrent items (r'_0) and for prices one-year lag (r'_{+1}), with the coefficient r'_0 larger. The coefficients for second and higher differences of the raw figures, and deviations as well, decrease as the order of difference increases; the coefficients for one-year lag of prices decreasing more rapidly than for concurrent items. These facts indicate that the maximum correlation of business cycles (including the irregular fluctuations) is for clearings preceding prices by less than half a year, say, four months. There is, however, an unknown element of spurious correlation between clearings and prices because the former are dependent upon prices as well as physical volume of goods exchanged and speculation. If this spurious element, due not to the method but to the nature of the data, were excluded, it is probable that the maximum correlation would be found for clearings preceding prices by more than six months, perhaps by a year.

The coefficients of correlation for the deviations all agree in locating the maximum, and therefore the lag of prices, at less than a year. The actual maximum found was for concurrent items, except for deviations from the nine-year averages which gives a maximum at one year lag of prices. Since our judgment is based upon the relative values of the coefficients for various degrees of lag, rather than upon their absolute values, the type of secular trend chosen does not appear to have great significance. Curve-fitting, however, does appear to be preferable to the taking of moving averages because, first, all the items may be used in determining the correlation and, second, the coefficients for deviations and first differences disagree in their location of the maximum when deviations from the mov-

ing average are taken but agree in all other cases. Of course it might be possible to use the deviations of all the terminal items from the moving average, if values of the latter were exterpolated; fitting a curve, graphically or otherwise, to the moving average is the obvious solution of this problem.

The results appearing in Table X make it clear that the coefficients for higher differences give little indication of the correspondence of the business cycles in the two series, which correspondence is clearly shown by the charts and the correlation coefficients for deviations from the secular trend and for first differences.

V.

If our interest were in the absolute degree of correlation between the cycles of two series for selected pairs of items, the nature of the curves used to represent the secular trends and the closeness of the fit to the data would be of primary importance. In case we are dealing with series in which the secular trends are non-linear, such as clearings and prices, but if, nevertheless, we use straight lines to represent the trends and correlate deviations therefrom, the coefficients resulting will undoubtedly contain a large spurious element, positive or negative. This is illustrated by the discrepancy between the coefficients, +.92 and +.73, obtained for deviations from straight lines and parabolas, respectively, of London clearings and Sauerbeck's prices (Table X). The former coefficient (+.92) undoubtedly contains a spurious element amounting to some twenty points. The spurious element is positive in this case, apparently, because of the downward long time trend from 1868 to 1896 and the upward trend from 1897 to 1913 which results in the pairing of large negative deviations for the period 1884–1899, when the deviations are from straight lines.

To test the effect of dividing the data of Sauerbeck's prices and London clearings into two homogeneous sub-periods, *i. e.* one of falling prices, 1868–1896, and one of rising prices, 1897–1913, two linear secular trends were found for each series and coefficients of correlation were computed for deviations from these trends and for their first differences. The lines

and their equations are given in Figure 9; the deviations appear in Figure 10; the coefficients are presented in Table XI. The maximum coefficients for the deviations are $r_0 = +.71$ and $r_{+1} = +.70$ for the periods 1868–1896 and 1897–1913, respectively. The maximum coefficients for corresponding first differences, $r'_0 = +.61$ and $r'_{+1} = +.47$, are consistent with those obtained for the entire period as shown by Table X. The coefficients for first differences, r'_0, between deviations from various trends (see A, B, C, D, and E of Table X) are +.42, +.42, +.55, +.49, and +.50.

TABLE XI.

COEFFICIENTS OF CORRELATION BETWEEN DEVIATIONS OF SAUERBECK'S PRICES AND OF LONDON CLEARINGS FROM THEIR RESPECTIVE LINEAR SECULAR TRENDS FOR THE TWO PERIODS 1868-1896 AND 1897-1913 TOGETHER WITH COEFFICIENTS FOR FIRST DIFFERENCES; VARIOUS DEGREES OF LAG OF PRICES (+) AND OF CLEARINGS (−).

Items Paired.	Coefficients of Correlation.				
	r_{-1}.	r_0.	r_{+1}.	r_{+2}.	r_{+3}.
			1868–1896.		
Deviations	+.38	+.71	+.63	+.34	−.07
First Differences	+.20	+.61	+.34	+.16	−.07
			1897–1913.		
Deviations	−.27	+.43	+.70	+.20	−.34
First Differences	−.40	+.27	+.47	+.04	−.40

Division of the data into two sections throws new light on the problem. Clearings and prices fluctuated concurrently during the first period, but prices lagged behind clearings by a year during the period 1896–1913. Perhaps increased speculation has changed the character of clearings during the second period. Whatever may be the cause, the fluctuations of English prices and clearings are shown to be related in the same fashion as are those for the United States during the same period (see AD, BD, and CD of Table IV).

TABLE XII.

COEFFICIENTS OF CORRELATION BETWEEN RELATIVE WHOLESALE PRICES AND PIG-IRON PRODUCTION OF THE UNITED STATES, DEVIATIONS AND FIRST DIFFERENCES AS FOLLOWS:

1879–1913.

A. Deviations from Linear Secular Trends and First Differences.

B. Deviations of Prices from Parabola and of Pig-iron Production from Compound Interest Curve, and Their First Differences.

1879–1896.

C. Deviations from Linear Secular Trends and Their First Differences.

1897–1913.

D. Deviations from Linear Secular Trends and Their First Differences.

1879–1896.

E. Deviations of Prices from Linear Secular Trend, and of Pig-iron Production from Compound Interest Curve (Computed for Data 1879–1913).

1897–1913.

F. Deviations of Prices from Linear Secular Trend, and of Pig-iron Production from Compound Interest Curve (Computed for Data 1879–1913).

1879–1913.

G. Deviations of Prices from the Two Linear Secular Trends (1879–1896 and 1897–1913) as a Continuous Series and of Pig-iron Production from Compound Interest Curve.

H. First Differences of Raw Items.

Prices concurrent (0), lag (+), or previous (−) to pig-iron production as indicated by subscript of r.

Symbol.	Period.	Items Paired.	Coefficients of Correlation. r_{-1}.	r_0.	r_{+1}.	r_{+2}.	r_{+3}.
(A)	1879–1913	Deviations (a) First Differences	..... +.31	+.76 +.41	+.74 +.21	+.63	+.62 .....
(B)	1879–1913	Deviations First Differences	+.05 −.37	+.48 +.40	+.51 +.22	....	....
(C)	1879–1896	Deviations First Differences	....	+.48 +.23	+.61 +.55	+.41 00	+.31 .....
(D)	1897–1913	Deviations First Differences	..... − 65	+.41 +.54	+.35 +.17	−.06	−.06 .....
(E)	1879–1896	Deviations	.....	+.54	+.69	+.49	+.20
(F)	1897–1913	Deviations	.....	+.55	+.48	+.10	+.11
(G)	1879–1913	Deviations	.....	+.45	+.44	+.19	+.10
(H)	1879–1913	First Differences	−.31	+.41	+.22	−.10	+.07

(a) Coefficient $r_{+4}=+.53$.

Table XII presents coefficients of correlation between deviations from various secular trends and their first differences of relative wholesale prices and pig-iron production of the United States for the period 1879–1913 and the sub-periods, 1879–1896 and 1897–1913. The various secular trends and their equations are given in Figures 11 and 12; the deviations appear in Figures 13 and 14; the data appear in Table XIII.

For the period 1879–1913 (see A, B, G, and H) the cycles of prices and pig-iron production are concurrent. For the period 1879–1896 (see C and E) the pig-iron cycles precede price cycles by a year. For the period 1897–1913 the cycles of the two series are strongly concurrent (see D and F). The coefficients of correlation for deviations from the linear trends, 1879–1896, $r_0 = +.48$ and $r_{+1} = +.61$ and $r_{+2} = +.41$, agree in locating the maximum at the same point as those for first differences, $r'_0 = +.23$ and $r'_{+1} = +.55$ and $r'_{+2} = .00$ (see C). The coefficients for deviations from the linear trends, 1897–1913, $r_0 = +.41$ and $r_{+1} = +.35$ are likewise supported by those for first differences, $r'_0 = +.54$ and $r'_{+1} = +.17$ (see D). Using deviations of prices from the two linear secular trends as *a continuous series* and of pig-iron production from the compound interest curve, 1879–1913, we have the coefficients $r_0 = +.45$ and $r_{+1} = +.44$ (see G). It is obvious that the coefficients for the whole period and the two sub-periods are consistent. *At present general prices and pig-iron production fluctuate concurrently.*

FIGURE 9.

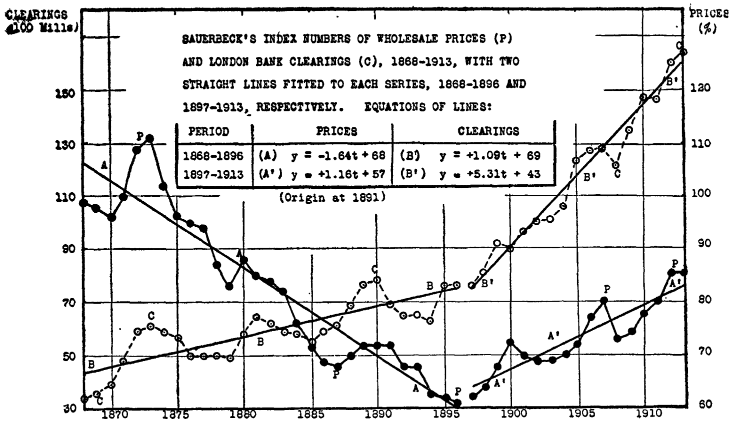
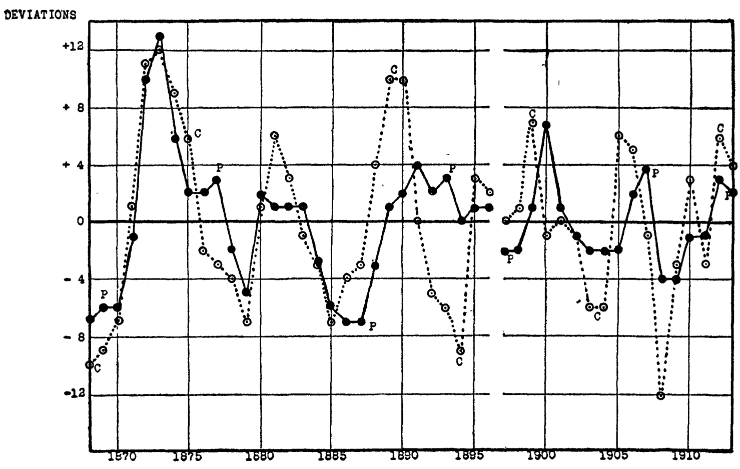


FIGURE 10.



DEVIATIONS OF LONDON CLEARINGS (C) AND SAUERBECK'S PRICE INDICES (P) FROM THEIR RESPECTIVE LINEAR SECULAR TRENDS FOR THE TWO PERIODS 1868-1896 AND 1897-1913.

FIGURE 11.

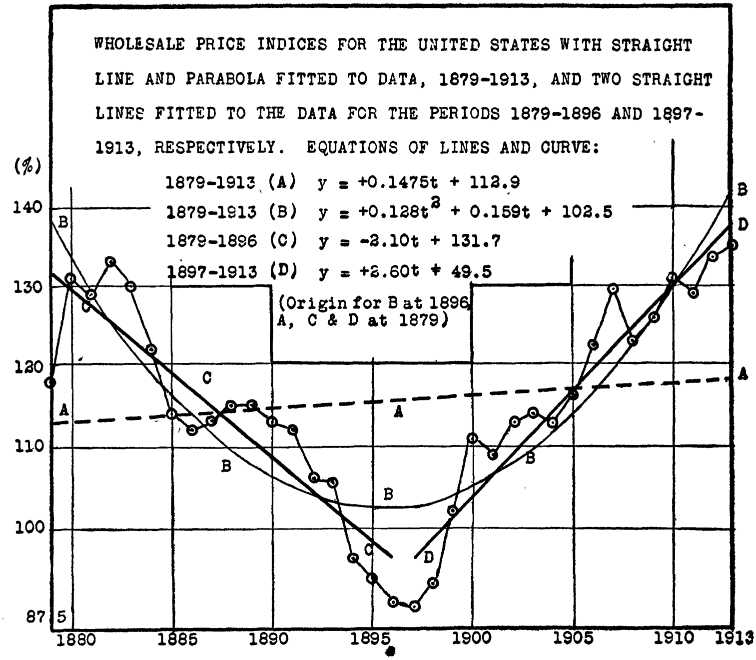


FIGURE 12.

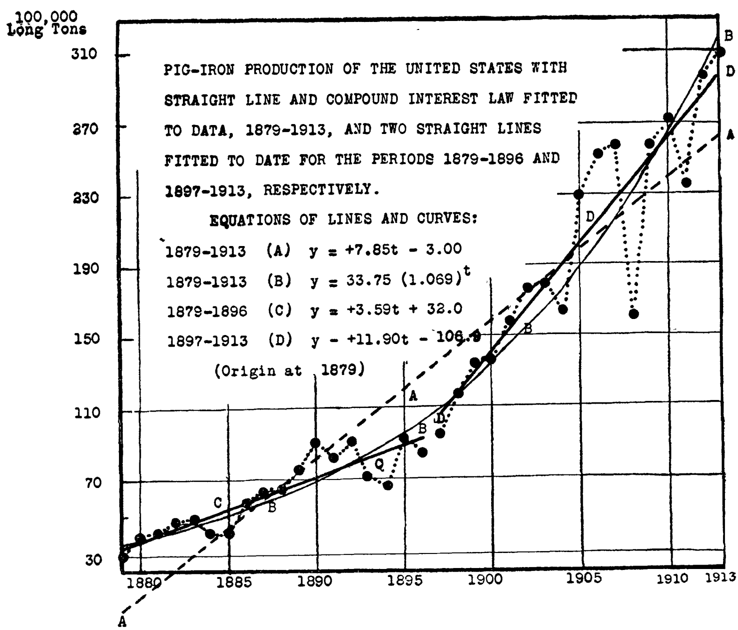
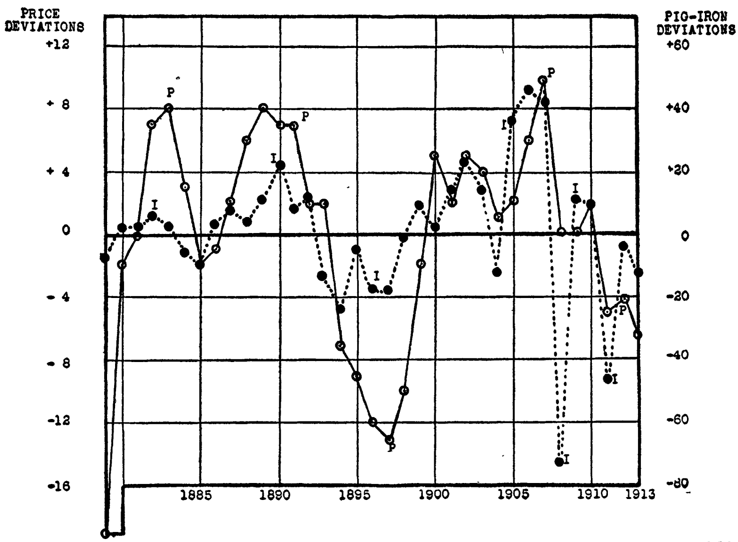
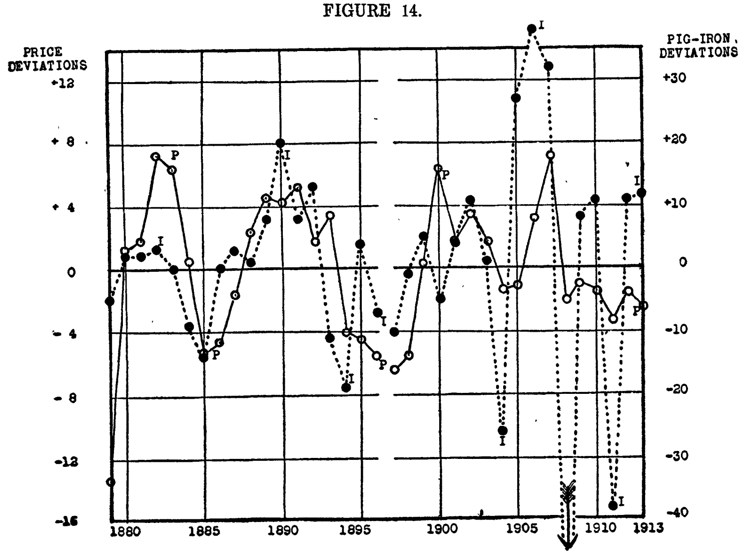


FIGURE 13.



DEVIATIONS OF UNITED STATES PRICE INDICES (P) FROM PARABOLA AND PIG-IRON PRODUCTION (I) FROM COMPOUND INTEREST CURVE, 1879-1913.

FIGURE 14.



DEVIATIONS OF UNITED STATES PRICE INDICES (P) AND PIG-IRON PRODUCTION (I) FROM THEIR RESPECTIVE LINEAR SECULAR TRENDS FOR THE TWO PERIODS, 1879-1896 AND 1897-1913.

TABLE XIII.

WHOLESALE PRICE INDICES FOR THE UNITED STATES (COLUMN 1) WITH THE DEVIATIONS FROM STRAIGHT LINE (2) AND PARABOLA (3) 1879-1913, AND FROM TWO STRAIGHT LINES, 1879-1896 AND 1897-1913 (4) AS SECULAR TRENDS. ALSO PIG-IRON PRODUCTION IN THE UNITED STATES (5) WITH THE DEVIATIONS FROM STRAIGHT LINE (6) AND FROM COMPOUND INTEREST LAW (7) 1879-1913 AND FROM TWO STRAIGHT LINES, 1879-1896 AND 1897-1913 (8) AS SECULAR TRENDS.

(Equations of Secular Trends in Figures 11 and 12.)

Year.	1.(a)	2.	3.	4.	5.(b)	6.(c)	7.	8.
1879	118.2	+ 5.3	−18.5	−13.5	27	+30	− 7	− 5
1880	130.8	+17.8	− 1.8	+ 1.2	38	+33	+ 2	+ 2
1881	129.3	+16.1	+ 0.4	+ 1.8	41	+28	+ 2	+ 2
1882	132.7	+19.4	+ 7.4	+ 7.3	46	+25	+ 5	+ 3
1883	129.7	+16.2	+ 7.7	+ 6.4	46	+18	+ 2	0
1884	121.6	+ 8.0	+ 2.6	+ 0.4	41	+ 5	− 6	− 9
1885	113.8	.0	− 2.5	− 5.3	40	− 4	−10	−14
1886	112.4	− 1.5	− 1.3	− 4.6	57	+ 5	+ 3	0
1887	113.3	− .8	+ 1.9	− 1.6	64	+ 4	+ 7	+ 3
1888	115.2	+ 1.0	+ 5.8	+ 2.4	65	− 3	+ 4	+ 1
1889	115.2	+ .8	+ 7.6	+ 4.5	76	0	+11	+ 8
1890	112.9	− 1.6	+ 6.8	+ 4.3	92	+ 9	+22	+20
1891	111.7	− 3.0	+ 6.8	+ 5.2	83	− 8	+ 8	+ 8
1892	106.1	− 8.7	+ 2.2	+ 1.7	92	− 7	+12	+13
1893	105.6	− 9.4	+ 2.4	+ 3.3	71	−36	−14	−11
1894	96.1	−19.0	− 6.6	− 4.1	67	−48	−24	−19
1895	93.6	−21.7	− 8.9	− 4.5	94	−29	− 4	+ 4
1896	90.4	−25.0	−12.1	− 5.6	86	−44	−18	− 7
1897	89.7	−25.9	−13.0	− 6.5	97	−41	−14	−10
1898	93.4	−22.3	− 9.9	− 5.5	118	−28	− 1	− 1
1899	101.7	−14.1	− 2.4	+ 0.2	136	−18	+ 9	+ 5
1900	110.5	− 5.5	+ 5.3	+ 6.4	138	−24	+ 2	− 5
1901	108.5	− 7.6	+ 2.0	+ 1.8	159	−11	+14	+ 4
1902	112.9	− 3.4	+ 4.8	+ 3.6	178	0	+23	+11
1903	113.6	− 2.8	+ 3.7	+ 1.7	180	− 5	+14	+ 1
1904	113.0	− 3.6	+ 1.1	− 1.5	165	−28	−12	−26
1905	115.9	− .8	+ 1.6	− 1.2	230	+29	+41	+27
1906	122.5	+ 5.6	+ 5.6	+ 2.8	253	+44	+51	+38
1907	129.5	+12.5	+ 9.7	+ 7.2	258	+41	+42	+32
1908	122.8	+ 5.6	+ 0.1	− 2.1	159	−66	−72	−79
1909	126.5	+ 9.2	+ 0.4	− 1.0	258	+25	+11	+ 8
1910	131.6	+14.1	+ 1.8	− 1.5	273	+33	+ 9	+11
1911	129.2	+11.6	− 4.5	− 3.5	236	−12	−46	−38
1912	133.6	+15.8	− 4.1	− 1.7	297	+41	− 4	+11
1913	135.2	+17.3	− 6.9	− 2.6	310	+46	−12	+12

(a) The Aldrich and Bureau of Labor Statistics indices are reduced to a continuous series with the base 1890–1899.
(b) From Statistical Abstract of the United States, 1914, p. 664. The units here are 100,000 long tons.
(c) Equation of line of secular trend, y=7.852t−3.00, origin at 1879. This is the only trend having any negative ordinate for the period studied.

The conclusion just stated is of especial interest because it is in conflict with that of Professor H. L. Moore.* In his *Economic Cycles*, he found the following coefficients between the cycles of crop yield and pig-iron production, *using three year averages in all cases:* $r_0=.625$; $r_{+1}=.719$; $r_{+2}=.718$; $r_{+3}=.697$; $r_{+4}=.572$ (see Table XIV). Correlating the cycles of crop yield with cycles of general prices,† he obtains the coeffi-

* Moore, H. L. *Economic Cycles*, p. 110.
† *Ibid.*, p. 122.

cients $r_{+3} = .786$, $r_{+4} = .800$, and $r_{+5} = .710$. He concludes from these coefficients, first, "that the cycles in the yield per acre of the crops are intimately related to the cycles in the activity of industry, and that it takes between one and two years for good and bad crops to produce the maximum effect upon the activity of the pig-iron industry" and, second, that "the cycles in the yield per acre of the crops are . . . . . . intimately connected with the cycles of general prices, and the lag in the cycles of general prices in approximately four years."* It seems to me that this conclusion is not warranted because of the poor fit of a linear secular trend to pig-iron production. The ordinate of the secular trend is *negative* for the years 1871, 1872, and 1873. The deviations from the secular trend are all positive for the periods 1871–1877 and 1902–1910 and all negative for the period 1878–1901. The deviations of crop yield are, with few exceptions, positive from 1871 to 1879 and 1903 and 1910 and negative from 1880 to 1902.† It appears probable, then, that the correlation coefficients upon which he relies contain a large spurious element. At any rate the *differences* between the coefficients, amounting to less than .02 in most cases, on which his judgment is based, cannot be considered significant.

Waiving the question of Moore's use of three-year progressive averages to form the series for which correlation coefficients are computed, which usage throws serious doubt upon the reliability of his conclusions, I will test the correlation between the series of three-year averages by computing the coefficients between first difference of those items. The coefficients are given in Table XIV. The coefficients between (1) crop-yield and pig-iron production and between (2) crop-yield and general prices are not significant. The former group coefficients (1) shows a curious alternation in value which, examination of the basic series demonstrates, is due to a few predominating items in pig-iron production after 1905. The latter group of coefficients (2) shows a maximum at four-years lag of prices but the coefficient ($r'_{+4} = +.39$) is not much larger than the maximum coefficient ($r'_{-2} = +.32$) found for first differences of the two random series previously analyzed (see

* Moore, H. L. *Economic Cycles*, pp. 110, 122.

† *Ibid.*, p. 131.

Table V). The coefficients for pig-iron production and general prices (3) reach a maximum ($r'_{+1} = +.59$) at one year lag of prices. This coefficient is probably significant. It gives, however, a result at variance with Moore's conclusions but consistent with the conclusions obtained when the period 1879–1913 was divided into two sub-periods (see Table XIII). Moore's use of three-year averages probably results in much higher coefficients than would result from annual figures. Even so, the coefficients between first differences are small for crops and the indices of the industrial and business activity. Moore's object was to show that the cycles of business reflect the cycles in crops, not merely, having assumed that cycles exist, to find the lag. For this object a good "fit" of secular trend to data is imperative. The method of first differences, then, is valuable because it reveals spurious correlation between deviations from secular trends when the fit is not good.

TABLE XIV.

(A) COEFFICIENTS OF CORRELATION BETWEEN DEVIATIONS OF YIELD PER ACRE OF NINE CROPS FROM THE LINEAR SECULAR TREND AND SIMILAR DEVIATIONS OF PIG-IRON PRODUCTION AND OF GENERAL PRICES, 1870–1911; ITEMS OF THE VARIOUS SERIES ARE THREE-YEAR PROGRESSIVE AVERAGES.

(B) COEFFICIENTS OF CORRELATION BETWEEN FIRST DIFFERENCES OF THE RESPECTIVE DEVIATIONS.

The subscript i in r_i indicates the lag in prices and pig-iron production compared with crops, or of prices as compared with pig-iron.

(A) Moore's Coefficients. (a)

Crop Yield Correlated with:	r_0.	r_{+1}.	r_{+2}.	r_{+3}.	r_{+4}.	r_{+5}.
Pig-Iron Production........................	.625	.719	.718	.697	.572	....
General Prices............................	....	....	....	.786	.800	.710

(B) Coefficients between First Differences.

Crop Yield Correlated with:	r'_0.	r'_{+1}.	r'_{+2}.	r'_{+3}.	r'_{+4}.	r'_{+5}.
(1) Pig-Iron Production..................	−.02	+.26	+.06	+.25	−.12	....
(2) General Prices........................	−.11	−.08	+.21	+.32	+.39	+.27
Pig-Iron Production Correlated with:	r'_{-1}.	r'_0.	r'_{+1}.	r'_{+2}.	r'_{+3}.	r'_{+4}.
(3) General Prices........................	+.25	+.51	+.59	+.44	+.25	....

(a) Moore, H. L. *Economic Cycles*, pp. 110–122. Series on pp. 131, 134.

Résumé.

The variate difference correlation method has been invented to eliminate spurious correlation due to position of items in time or space.

The method involves the assumption that the taking of multiple differences leads to series of random variates. In practice for short series this assumption is not fulfilled.

Coefficients for higher differences of short series tend to alternate in sign and to conceal rather than to reveal the nature of the correlation between the series being tested.

Stability of coefficients for higher differences appears to have little significance for short series, and perhaps for long series as well. The assumption that the series correlated are made up of variates "randomly distributed in time," if fulfilled, will lead to stable coefficients for successive differences. However, though this condition is *sufficient* for stability it is not *necessary*.

In testing economic series for correspondence of their cyclical fluctuations, especially in determining the relative position of the cycles upon the assumption that there are cycles, the correlation coefficients between deviations from a linear secular trend together with coefficients for first differences constitute a reliable basis for judgment.

When the question is one of the *existence* or *non-existence* of similar cycles in two time series great care must be used in the choice of the function used to represent the secular trend and in the nature of the fit of the curve or line to the data. The method of first differences is an extremely valuable aid in investigating such a question.

Coefficients of correlation between second differences may give information concerning minor oscillations as distinct from secular trend and major cycles. Even for this purpose the use of higher than second differences appears to be unreliable, especially so for short series. The coefficients of correlation between second differences are identical with those between deviations from three-year progressive averages.

The method of measuring correlation between cycles of time series, that is both easy of application and reliable, is the method of *first* differences. In general, however, this method

should be supplemented by curve fitting. To secure a picture of the cycles, it is, of course, necessary to take deviations from a closely fitted curve.

Finally, curve fitting to eliminate the secular trend of a time series should always be adapted to the problem in hand and interpretation of coefficients of correlation between time series should be made with continual reference to the fundamental data. Important light may be secured by dividing statistical series into more homogeneous sub-series and analyzing the latter. The nature of the data is as important as the method to be applied. Rules-of-thumb concerning method or data are apt to lead to pitfalls.